Potentially Paltry Pipeline Ponderings with Perhaps a Few Proposed and Possibly Piddling Pertinent Peregrinations: a Partial Presentation

Marc Murison

U.S. Naval Observatory
Washington, DC

February 15, 2001

Introduction

Reduction of FAME data presents a challenge that is most likely surmountable. Even at the broadest level, it is not clear how to successfully construct and then sculpture a robust data reduction scheme. Currently it is envisioned that we will copy the Hipparcos architecture, namely that, at least conceptually, there will be three segments of the reduction pipeline: "spiral" reduction, global reduction, and astrometric parameter estimation. Besides being intuitively plausible, this approach has two features that recommend it: through Hipparcos it has an existence proof of sorts (the actual correspondence is tenuous at best), and it allows the skeleton FAME team to avoid creative hard thinking. (The latter would, among other things, require time and staffing that the FAME project has never been given). Due to FAME employment of CCD detectors, a *de facto* additional reduction stage will involve PSF centroiding prior to "spiral" reduction. See the data reduction data flow diagram at the FAME web site for an overview.

Each of the "proposed" (mainly through lack of a plausible alternative approach) four reduction stages involves at its heart an iterative, nonlinear least squares parameter estimation problem. The dependencies of the parameters most relevant to each stage upon the parameters and peculiarities of the other stages argues for a somewhat holistic approach, but the different timescales inherent to each stage and the existence of parameter subsets that are more isolated to one stage or another argue for a certain degree of reduction independence between the stages. This is the basic conflict that is currently a contributor in part to the current confounding of formative attempts at construction of an overall data reduction approach.

This fundamental conflict will not be solved, discussed, or even addressed here. Instead, in accordance with FAME tradition, I will don blinders and consider only a few very limited aspects of just the so-called "spiral" reduction stage. (They're not spirals!) I will attempt (and largely fail) to raise questions regarding the problems we may face, what we might need, and perhaps how we might think about tackling those problems. It goes almost without saying that adequate time has not been devoted to do either a good or a complete job of it, but it's a start.

Nonlinear Least Squares is the Least of Our Worries

Fortunately, parameter estimation via weighted nonlinear least squares (WNLS) is a fairly mature field. Hence, coupled with the fact that the FAME data set size is not remarkable when compared to truly high data volume projects in various fields (solar physics, high-energy physics, geophysics, etc.), this means there is no need for the FAME team to invent anything new in this area. All currently-envisioned FAME parameters are thought to exhibit characteristics that render the data set amenable to well-tested WNLS algorithms:

- smooth temporal behavior during the relevant timescale (~several hours for "spiral" reduction)
- behavior is well-approximated either via a physical model or via a bias model or both
- hence, well-defined analytical derivatives

Ideally, anyway.

WNLS: Here's How it Works

We start with the usual weighted cost function, consisting of summed squares of O-Cs, or "data minus model" values. We safely assume that the model functions are nonlinear with respect to the parameters that need to be estimated — hence the need for a nonlinear algorithm, invariably iterative in nature. Consider the cost function to second order near the "best" parameter values — i.e., the "parabolic" approximation:

$$\chi^2(\boldsymbol{a}) \approx \chi^2(\boldsymbol{a}_0) - \Delta\boldsymbol{a}\cdot\boldsymbol{B} + \tfrac{1}{2}\Delta\boldsymbol{a}\cdot\boldsymbol{M}\cdot\Delta\boldsymbol{a}$$

where $\boldsymbol{a}$ is the vector of parameters, $\boldsymbol{a_0}$ is the set of "best" values that we hope to determine,

$$\boldsymbol{a} \equiv \boldsymbol{a}_0 + \Delta\boldsymbol{a} \; ,$$

and we must form the derivatives

$$\boldsymbol{B} \equiv -\frac{\partial\chi^2}{\partial\boldsymbol{a}}, \qquad [M]_{ij} \equiv \frac{\partial\chi^2}{\partial a_i \partial a_j}$$

from the model functions. The correction vector $\Delta\boldsymbol{a}$ which minimizes $\chi^2(\boldsymbol{a})$ is given by the value for which the parameter correction gradient vanishes:

$$\frac{\partial\chi^2(\boldsymbol{a})}{\partial\Delta\boldsymbol{a}} \approx -\boldsymbol{B} + \Delta\boldsymbol{a}\cdot\boldsymbol{M} = 0$$

Hence, starting with *a priori* parameter values (as an iteration starting point), we calculate the parameter corrections

$$\Delta\boldsymbol{a} = \boldsymbol{M}^{-1}\cdot\boldsymbol{B}$$

Using the updated parameter values, we calculate a new set of corrections, and so on until the corrections hit the noise floor (i.e., the parameter values stop changing significantly). There are various methods of accelerating convergence and/or increasing solution robustness (Levenberg-Marquardt is popular), and inversion of large matrices (with various properties) is a field unto itself (also mature, fortunately), but this is the basic algorithm.

WNLS: Potential Problems

The inverse of the *curvature matrix* ***M*** is the *covariance matrix.* It contains information not only about the formal parameter estimation errors (the diagonal elements) but also about the parameter cross correlations (the off-diagonal elements).

Already, we can identify at least five potential problem areas with respect to the FAME "spiral" reduction parameter determination (in no particular order):

1. **Large parameter cross correlations.** If the errors of two or more parameters are correlated, then the model cannot formally discriminate between those parameters, and we therefore cannot determine to which parameter the associated error really belongs. This also represents a dilution of information, leading to a degredation of errors throughout the parameter vector and ultimately of the errors in the Holy Grail (astrometric parameters). Parameters masking other parameters is inevitable in a system as complex as this. Dealing with it is somewhat of a black art.

2. **Coupling between the large-scale reduction stages.** The intersections of parameter sets relevant to the large-scale segments of the data pipeline (centroiding, "spiral", global, astrometric) will be nonzero. Currently, we don't know the extents of overlap, but we can safely assume they will be significant. For example: the basic angle behavior directly affects both the centroiding and "spiral" stages. Thus, we must concoct an iterative mechanism whereby the couplings can be taken into account and the various parameter intersection sets can be further refined by using information available from the "other" reduction stages — without driving the algorithmic complexity into the stratosphere or the numerical robustness down the sewer. But I said I wasn't going to discuss this here, so I shall formally cop out now.

WNLS: Potential Problems *(continued)*

3. **Number of parameters.** Inversion of the curvature matrix may be a problem due to the large number of parameters versus the limited amount of data per contiguous data segment ("spiral" length). This issue might possibly hinge on how detailed the characterization of the CCDs must be. If we must determine parameters for each CCD pixel, for example...(not a happy thought).

4. **Data discontinuities.** Discontinuous behavior of the data comes in two forms: data dropouts (noncontiguous data segments) and derivative discontinuities (contiguous but only piecewise continuous data segments).

   Dropouts occur due to a variety of causes, such as electronics blurps, waiting for thermal shock vibrations to damp to reasonable amplitudes, ground operators taking naps, etc. If typical intervals between dropouts are of order a few spin periods or greater, then dropouts do not pose a problem for the "spiral" reductions. However, this may be a big if, so we should prepare to handle dropouts *within* "spiral" segments.

   Piecewise continuous behavior is potentially more serious. This can occur when sudden changes occur within the instrument. Shifting of optical elements and microquaking of structural members and optical materials (esp. the primary and compound mirrors) during thermal variation cycles are examples. (When dealing in microarcsecond realms, the kinds of effects that suddenly become important can be surprising.) When derivative breaks occur, the normally well-behaved model fit sails right out the window. Kalman filters may be the answer for this problem — though, as far as specific knowledge within the overextended skeleton FAME team is concerned, this is just hearsay and must therefore be investigated.

5. **Ill-conditioned curvature matrix.** The curvature matrix is likely to be ill-conditioned, at least for some data segments. This actually belongs in the category of matrix inversion, so we can probably assume there is a variety of tricks already available both for dealing with it and for interpretation. It does mean, though, that some parameter values will at least sometimes be flaky.

"Spiral" Reduction Parameters

It is most efficient, from an information theory perspective, to estimate as few parameters as possible. This distributes the information contained in the data so as to have maximum impact (i.e., minimizes the parameter errors), at least on average. (The actual distribution of information among the parameter determinations is, of course, a much more difficult assessment.) To that end, then, each physical process (for example, basic angle variation, s/c attitude, etc.) might be modeled in the form

$$y(t,\boldsymbol{a},\boldsymbol{b}) = \varphi(t,\boldsymbol{a}) + \sum_{k=1}^{n} f_k(t, b_k)$$

where $\boldsymbol{y}$ is the model, ϕ is a function that contains all the relevant physics to high enough order to reach the systematic error floor, and the summation term is a bias model that attempts to account for systematic errors (which in turn arise from inadequacies of the phsycal model ϕ). The bias model is composed of appropriate basis functions $\boldsymbol{f}$ that depend on the bias parameters $\boldsymbol{b}$. Ideally, $\boldsymbol{n}$ is very small.

What kinds of parameters are we likely to need to estimate during the "spiral" reduction stage? This is a harder question than it seems at first. Three broad categories immediately suggest themselves.

1. **Target characteristics.** This category of parameters (including stellar temperature, luminosity, metallicity; reddening; multiplicity; etc.) comes to the "spiral" stage indirectly from the centroiding stage. Systematic errors in the determination of these parameters during the centroiding stage result in systematic errors in the primary centroiding stage product, the target centroid determinations. These in turn will give rise to a systematic error component of the primary "spiral" stage parameters — those of the attitude model — as well as of the "spiral" stage contributions to the instrument parameter estimations. Gnarly.

"Spiral" Reduction Parameters *(continued)*

2. **Instrument parameters.** This category includes all the parameters associated with instrument behavior: orientations, positions, and surface figures of all optical elements (including the curved focal plane window); and orientations and positions of the CCDs (wrt the focal plane assembly), the focal plane assembly (wrt the instrument), and the instrument (wrt the s/c bus). The spacecraft thermal environment and its behavior with time directly affect all of these parameters. That is, these parameters are functions of the thermal behavior. The thermal behavior in turn has both a time dependence and a spatial dependence. Gnarly.

3. **Attitude model parameters.** The spacecraft attitude model is the primary product of the "spiral" reduction stage. The perturbation environment is somewhat complex — see the web page http://aa.usno.navy.mil/SymTop/SpinDynamicsIssues.html for a partial list — but there is well-known physics behind almost all of it.

   Perturbation amplitudes will be determined by various spacecraft parameters such as mass distribution, spacecraft geometry, outer surface coefficients (i.e., radiation reflectance, absorption, and diffusion coefficients), etc. Phases will also need to be determined. Many of the various perturbations have the same periods (e.g., precession period, orbital period, multiples of the spin period, etc.), so for effects that cause similar-amplitude perturbations of the attitude there will be a parameter masking issue. It may be possible to combine some perturbations into an *ad hoc* model with fewer parameters that achieves the same (or similar enough) accuracy — we should stay open to that possibility.

   In general, though, we have a bunch of physical parameters to estimate, and no doubt a set of bias model parameters as well. How many of these parameters there are is currently an open question (the number of bias terms needed can be determined in simulations), but I suspect that it is few compared to the number of observations in most data segments. So the solution will converge, and the real question is the magnitudes of the parameter estimation errors. This can be addressed by covariance studies, though the appropriate time for those was of order 2-3 years

"Spiral" Reduction Parameters *(continued)*

ago.

The most problematic classes of perturbations are the stochastic and hard-to-model effects. This includes effects such as the solar irradiance fluctuations, Earth weather, and fuel sloshing. By a tremendous stroke of luck, the two viewports are nearly orthogonal. Hence we can use grid star cross-scan measurements in one port as a useful constraint on the attitude for measurements in the other port (see http://aa.usno.navy.mil/SymTop/TimingErrors.html). Nevertheless, stochastic perturbations are going to lead to errors in the attitude model, and we will most likely need to devise an approach for mitigating these kinds of errors.

I should mention before I forget that the attitude parameter estimation process will, among other things, involve integration of the dynamical equations of motion (spin and orbital). Initially, parameter values will be assumed (including dynamical state vector initial conditions), the model integrated forwards in time over the length of a particular data segment, and the residuals formed from comparison of the integration to the data. Based on the mismatch to the data, parameter corrections will be determined (see page 4 of this document), new parameter values will be calculated, and we do it over and over again until we have parameter values that cause the model to most closely match that data segment. This is just the WNLS algorithm previously mentioned, but the point is that numerical integration of the dynamics is part of that process. Experience shows that only a few iterations (<10, often <5) are needed for convergence to the systematic error floor.